5

TECHNIQUES FOR INVENTORY VALIDATION

5.1 UNCERTAINTY ISSUES

Much of the discussion in this report addresses the current status and remaining weaknesses associated with the tools and procedures available for estimating emissions of PM-2.5 and its precursors. In applications to PM-2.5 planning, as well as for all other air quality planning programs, the emissions inventory development techniques rely primarily on estimation methods. Emissions data based on actual source measurements for PM-2.5 planning will be limited. In most cases, average emission factors are applied to all sources in a particular process or source category, and, in many cases, default values or procedures are used to estimate the activity data to which the emission factor is applied. In practice, the process characteristics and operating rates of individual sources can vary considerably within any particular source category. In addition, uncertainties arise through the use of surrogate spatial, temporal, and speciation allocation files. Therefore, emission inventory developers and the team responsible for air quality planning efforts must be aware of the inherent uncertainty of the underlying emissions estimates. Those uncertainties must be considered while establishing priorities, identifying causes of observed air quality, and ultimately in deriving control programs to achieve air quality goals.

The concept of uncertainty as applied to emissions inventory development is fundamentally different than the quantitative uncertainty and error bounds that are applied to measurements in experimental programs. The process to estimate and interpret uncertainty in emissions inventory applications is more intuitive than the approach that is commonly applied in experimental research. Therefore, the developers of PM-2.5 emissions inventories and the planners that depend on those inventories must be cautious when interpreting results based on these emissions estimates. Programs must maintain the flexibility to adjust to shortfalls and weaknesses of programs that are based on these inherently uncertain emissions estimates.

Because PM-2.5 planning is just getting started, the uncertainties associated with PM-2.5 emissions estimates can have more serious effects than in the case of other more mature planning programs. The techniques and tools used to estimate emissions inventories for PM-2.5 programs and to interpret analyses based on those inventories will improve as

experience is gained. Specifically, EPA and other organizations like the EIIP are working to improve emission factors, procedures to collect activity data, chemical species allocation profiles, and spatial and temporal allocation factors. Additional improvements in methodologies and tools for use in preparing emission inventories for PM-2.5 will also be developed by State planning agencies. During the initial planning efforts, it will be necessary to review the emissions estimates and to evaluate how well those estimates represent the overall air quality problems that are being addressed. The remainder of this section provides some suggestions for activities that can be used to validate emissions data and to identify components of the inventory that do not support other observations.

The following discussion presents a selection of techniques that States might find useful to complete this type of evaluation. The techniques discussed here are not the only methods that can be used to validate emissions inventories, but are discussed as examples of the kind of analyses that can be applied. States are encouraged to develop additional methods and techniques to assist in the evaluation of inventory data quality.

5.2 SPECIATED LINEAR ROLLBACK AND SIMPLE REGRESSION MODELS

The concept of linear rollback was first applied in early efforts by EPA to understand and prioritize the relative effects of various emissions sources on ambient air quality problems. This technique is based on the assumption that reductions in the amount of emissions from sources and/or source categories that affect a particular air quality problem will result in proportional reductions in the air quality measurement. This assumption does have some validity, although most air quality problems, including PM-2.5, do not exhibit a one-to-one relationship between emissions and air quality.

The speciated linear rollback model is based on the assumption that the mass of each type of particle (e.g., sulfate, nitrate, elemental carbon, etc.) is related to the spatially averaged emissions of the relevant pollutants. The requirements for the application of this approach are a speciated inventory covering the major contributors within the region, and a speciated ambient data set that represents the temporal averaging of the primary sources. The model will identify the contributing sources or source categories based on the resulting ambient air quality. The model can be used to complete preliminary tests of the adequacy of the regional inventory and to help prioritize efforts to find sources that may not be adequately represented or those that may not be represented at all. In this regard, it is a simplified receptor model.

Regression models use empirical relationships derived from the source and ambient data to apportion ambient particulate data among the distribution of sources in the region. The technique requires that data are available over a range of conditions and over a reasonable expanse of time. If sufficient information is available, these models can attribute source

contributions for the precursors to secondary particulate, as well as, the direct emissions of primary particulate.

Both of these techniques can be applied with limited information in a screening mode to identify serious weaknesses in the underlying emissions data. Take as an example a case in which these techniques indicate that there must be a major source or collection of sources that contribute organic carbon. Upon review of the inventory, no sources of organic carbon with a sufficient magnitude are found. This result would lead to an investigation of the possible sources of the missing organic carbon. Details of the application of the speciated linear rollback model and regression models can be found elsewhere. (NRC, 1993)

5.3 DATA ATTRIBUTE RATING SYSTEM (DARS)

The Data Attribute Rating System (DARS) was developed to assist in evaluating data associated with emission inventories. DARS provides a numerical confidence rating for emissions inventories. The numerical result is called the DARS score and in most cases it is less subjective than the usual qualitative letter grade rating procedures (i.e., A through E) that have historically been used to characterize emissions data quality. DARS was originally developed as a research tool for rating national and global greenhouse gas inventories. State agency personnel have used DARS to rate their base-year State Implementation Plan (SIP) ozone precursor inventories. In addition, particulate matter (PM-10) inventories (State- and National-levels) were evaluated by inventory developers trained in the use of DARS. More detail on the development and application of DARS can be found in EIIP documents available on the EIIP Internet site shown previously. (EIIP, 1996)

The DARS score is based on the perceived quality of both the emission factor and activity data. Numerical scores are assigned to four data attributes: measurement/method, source specificity, spatial congruity, and temporal congruity. These scores, which range from 1 to 10, reflect the confidence that the user associates with each of the attributes. DARS scores can be applied to groups of sources rather than to individual sources or to emissions estimates aggregated at different spatial and temporal scales. Results of these types of analyses are often useful to help assess priorities by defining large emissions sources that have low confidence levels. One advantage of DARS is to provide a quick evaluation of the effect of National-level or surrogate factors and activity data relative to local source specific factors. The relative improvement in the total score will be indicative of how sensitive the emissions are to local influences.

The proposed applications of DARS include:

- to validate emissions estimates to identify the weakest areas of an existing inventory for further research and improvement,
- to quickly compare and rank different inventories,
- to rank alternative emission estimation methods, and
- to set Data Quality Objective (DQO) targets during inventory planning and for future inventories.

5.4 USE OF CMB AND DISPERSION MODELING

Dispersion and receptor modeling are powerful tools for identifying and assessing various sources of primary emissions. The utility of these tools increases for applications to chemically inert species, but under some conditions, these tools can be used to evaluate sources of reactive components as well. One receptor modeling approach based on the Chemical Mass Balance Model (CMB), has been used effectively in many PM-10 planning efforts. The results of CMB modeling often can point to the major sources of the primary particulate and even to give estimates of relative source magnitudes. Similarly, dispersion modeling of chemically inert components can provide important information on the relative accuracy of emissions estimates. In both of these applications ambient concentration data are required.

While total mass concentration can be applied to some analyses, the power of these tools increases dramatically when applied to data that represent a full range of chemical speciation in the ambient data set. Currently, there are no ambient data for PM-2.5 collected using the FRM. Monitors consistent with the FRM are being deployed, and initial data from the first of these monitors will be available in 1999. The current monitoring plan requires that a subset of these monitors will collect samples for analyses of the chemical composition. These data will provide a wealth of information to use in CMB and dispersion modeling analyses to verify important sources, evaluate emissions estimates, and to prioritize emissions inventory improvement efforts.

Of course, the use of these approaches also requires detailed source composition data. Currently, the bulk of PM-2.5 source speciation data available through the SPECIATE system are dated, and in many cases, are not source specific. The only source-specific chemically speciated source data for PM-2.5 have been developed in research field studies. Some of these data in both SPECIATE and through the research programs include the condensable fraction and others do not. Many of the studies were completed prior to the promulgation of the NAAQS with the specification of the 2.5 µm size cutoff; therefore, some available profiles may be based on other size cutoff limits. Even with the difficulties presented with the existing data, it may still be possible to apply either CMB or dispersion

modeling to help States evaluate emissions estimates for the specific sources and source combinations that are important in particular areas.

This type of analysis will be most useful when applied across seasons, emissions scenarios, and other temporally variable scales that can affect the emissions from specific sources or the mix of emissions from the collection of sources in any given area. In the majority of cases, emissions data are expected to be applied to analyses of annual ambient concentrations. The causes of high ambient concentrations might change, however, with season or other temporally cyclical conditions. Repeated comparisons of the composition of the ambient samples to source components can be extremely informative in terms of improving the emissions inventory, identifying the principal causes of high measured concentration, and ultimately, in developing effective control plans.

5.5 DEVELOPMENT OF SOURCE PROFILES FOR CMB MODELING

Most local sources of primary PM-2.5 will be combustion sources, and the emissions will include a significant fraction of carbon. Direct sampling data and/or estimates of emissions from sources that are similar to other measured sources will benefit significantly if the elemental and organic carbon fractions are differentiated. This is such a critical requirement that States are cautioned that sampling of sources is not expected to be cost effective if results are not available for both elemental and organic carbon. Speciation of other components of the source mix that can serve as unique markers for specific sources will also improve source apportionment analyses. Developing the source profiles can be an expensive part of any source apportionment project, and the testing phase should be planned carefully to develop all of the data necessary.

Other issues of concern in developing source profiles are listed below:

- Ensure that plans are made to collect data with sufficient chemical speciation to allow CMB to distinguish specific source contributions.
- Develop profiles for all of the important sources affecting the receptor. Keep in mind that profiles that are specific for the actual sources of interest improve the confidence in the results relative to analyses that use source average or surrogate profiles.
- Complete screening tests to determine if two important sources have characteristics that are so similar that CMB cannot distinguish between them. If that appears to be the case, consider adding a chemical marker to one or both of the sources to make each source unique.
- Test the sources under the typical operating conditions that are believed to result in the air quality effect.

- Be sure to collect data that is representative of the spatial scale and the temporal scenario under which the air quality effect is observed.